

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

A spatial stochastic frontier model introducing inefficiency spillovers

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Galli, F. (2023). A spatial stochastic frontier model introducing inefficiency spillovers. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS, 72(2 (May)), 346-367 [10.1093/jrsssc/qlad012].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/918308> since: 2023-02-27

*Published:*

DOI: <http://doi.org/10.1093/jrsssc/qlad012>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**Federica Galli. (2023). "A spatial stochastic frontier model introducing inefficiency spillovers". *Journal of the Royal Statistical Society Series C: Applied Statistics*, Vol. 72, Issue No. 2, May 2023, Pages 346-367.**

The final published version is available online at:

<https://doi.org/10.1093/jrsssc/qlad012>

#### Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# A Spatial Stochastic Frontier Model Introducing Inefficiency Spillovers

Federica Galli

*University of Bologna, Bologna, Italy.*

E-mail: federica.galli14@unibo.it

**Abstract.** This paper develops a spatial Durbin stochastic frontier model for panel data introducing spillover effects in the determinants of technical efficiency (SDF-STE). The model nests several existing spatial and non-spatial stochastic frontier specifications and is estimated using maximum likelihood techniques. Estimates are shown to be unbiased even for small sample sizes and for alternative specifications of the spatial weight matrix implementing different Monte Carlo simulations. Finally, an application to the Italian accommodation sector is provided. Empirical findings suggest the relevance of the SDF-STE model in capturing labour productivity and knowledge spillover effects.

**Keywords:** Spatial stochastic frontier models, Spillover effects, Firms' efficiency, Italian accommodation sector.

## 1. Introduction

In the last two decades, it has been widely demonstrated that biased estimates and invalid statistical inference can occur in evaluating the level of productivity of firms using stochastic frontier models (SF) if spatial correlation is not taken into account (Areal and Pede 2021; Fusco and Vidoli 2013; Glass, Kenjegalieva, and Sickles 2016). Indeed, firms tend to cluster toward similar locations to take advantage of geographical proximity thanks to emulation, an extensive local market for specialized inputs, and shared knowledge. Therefore, to consider spatial dependence and to capture global and local spillovers, authors began to extend classical SF models in order to include also spatial terms.

Spatial dependence is usually introduced in SF models through the spatial lag of the dependent variable (SAR term) or through the spatial lag of the production inputs (SLX term). Specifically, the SAR term enables to capture productivity spillovers while the SLX term represents input spillovers, where the former refers to the attempt of less efficient producers to emulate the industry leader's best practices in order to increase their productive capacity (Syverson 2011), while the latter concerns the possibility that an extensive local market for workers and specialized inputs can allow clustered firms to better combine the productive factors reaching higher productivity levels (Porter 1998). Besides productivity and

input spillovers, in economic geography literature, knowledge spillovers, meaning “*working on similar things and hence benefiting much from each other’s research*” (Griliches 1992, p. 29), are usually acknowledged as the third kind of spatial effects. Indeed, knowledge and informational spillovers can increase firms’ performance more than for isolated producers because the geographical concentration of firms stimulates their innovative activity, spreading new knowledge through a tacit diffusion process (Hoover 1948).

Despite the general consensus of economists about the relevant and positive role of knowledge spillovers in affecting clustered firms’ industrial activity (Adams and Jaffe 1996; Griffith, Harrison, and van Reenen 2006; Levin and Reiss 1988; Spence 1984), to our knowledge, they have not yet been introduced in spatial stochastic frontier models. Specifically, firms’ innovative activity can be considered as one of the main determinants of firms’ efficiency level and therefore, knowledge spillovers can be identified as spatial effects arising from the factors that determine neighbouring firms’ efficiency. Therefore, we introduce the possibility to evaluate whether the determinants of technical inefficiency (hereinafter, denoted as  $Z$  variables) of nearby firms contribute to shaping the efficiency level of neighbours. To achieve this goal, we include the exogenous spatial lag of the  $Z$  variables in the inefficiency model in the same fashion as a standard SLX model. Besides considering spillover effects in the inefficiency determinants, we also include global and local spatial effects related to the frontier function following Glass, Kenjegalieva, and Sickles (2016). The resulting model (SDF-STE) enables to consider different sources of spatial dependence using a spatial Durbin specification for the frontier function and introducing for the first time the spatial lag of the inefficiency determinants.

Thus, we propose a novel SF model for panel data which includes three spatial terms capturing global productivity spillovers through the SAR term, local input spillovers through the SLX term and determinants of inefficiency spillover effects adding a new spatial term, that is the spatial lag of the  $Z$  variables. In particular, we differentiate between spillover effects influencing firms’ productivity and efficiency levels. Specifically, while both global and local spatial spillovers are considered for the frontier function following Glass, Kenjegalieva, and Sickles (2016), we concentrate on local spillover effects associated with the determinants of firms’ inefficiency because spatial dependence influencing neighbouring firms’ efficiency level mainly arises from local factors such as emulation, face-to-face interactions, local cooperation, and individuals contact (Griliches 1992).

The SDF-STE model can be estimated using a ML estimation approach. Moreover, direct, indirect and total effects affecting firms’ productivity and efficiency levels respectively originating from the productive inputs and from the determinants of firms’ efficiency can be computed following the method proposed by LeSage and Pace (2009). This is the first model that allows evaluating the role of spatial spillovers in affecting the efficiency level of neighbouring producers. In particular, the first new feature of our model consists in introducing the possibil-

ity of directly evaluating how each variable that determines the inefficiency level of neighbours also affects nearby firms. The second relevant feature of our model concerns his general and comprehensive specification enabling to capture different kinds of spatial spillovers across firms (i.e. productivity and input spillovers affecting the frontier function and local spillovers in the  $Z$  variables). Moreover, this model nests several existing spatial and non-spatial specifications, enabling to select the model that best fits the data testing different restrictions through likelihood ratio tests.

From an applied perspective, the proposed spatial specification taking advantage of new lagged variables can provide policymakers with interesting and relevant policy implications that could have not been obtained using previous spatial SF models. Indeed, unlike the previous spatial SF models that allow evaluating the overall level of spatial dependence affecting firms' inefficiency level by providing only generic information, the SDF-STE model makes it possible to evaluate the specific spatial effects arising from each inefficiency determinant giving rise to precise, detailed and distinct insights concerning spatial spillovers related to each variable of interest. This approach can be very useful for a variety of economic problems dealing with agglomeration economies, knowledge and *R&D* spillovers, technology diffusion, imitation, spatial networks and interactions, which are all relevant and current research topics in firm-level microdata applications, in productivity and efficiency analysis, in business-strategy literature, and in the context of regional sciences.

The rest of the paper is organized as follows: Section 2 discusses previous spatial SF models; Section 3 introduces the specification of the SDF-STE model, the estimation method and the derivation of the marginal effects and of the TE scores; Section 4 provides some Monte Carlo experiments testing the final sample properties of the estimated parameters; Section 5 shows an application to the Italian accommodation sector and Section 6 presents the final remarks.

## 2. Literature Review

Firms' productive performance is commonly analyzed in empirical contexts using SF models, or alternatively, through the non-parametric data envelopment analysis (DEA). The main reason for choosing SF models over standard DEA is that the latter is extremely sensitive to outliers and it does not allow for statistical inference. On the other hand, by estimating SF models using ML techniques, it is possible to formulate different hypotheses on the efficiency error term and on all the other estimated parameters implementing standard Wald tests, likelihood ratio tests or Vuong tests based on the Kullback-Leibler information criterion for non-nested hypotheses (Coelli 1995; Lai and Huang 2010). The classical specification of SF models was firstly introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van Den Broeck (1977) splitting the error term into two independent components,  $v_{it}$  and  $u_{it}$ , representing the random disturbance and

the inefficiency error term, respectively distributed as a normal and a truncated normal random variable. Subsequently, Battese and Coelli (1995), proposed to model the mean of the technical inefficiency error term by introducing some determinants of technical inefficiency for each firm at each time period.

Starting from these basic SF models assuming cross-sectional independence, in the last two decades, many authors began to claim that the assumption of spatial independence is not appropriate. Indeed, firms tend to cluster and share information and they cannot be considered isolated entities. Therefore, firms' location and interactions among nearby producers should be taken into account. In particular, as highlighted by Schmidt et al. (2009) and Glass, Kenjegalieva, and Sickles (2016), the omission of the spatial lag of the dependent variable capturing global productivity spillovers can result in biased estimates, due to an omitted variable bias. Moreover, Fusco and Vidoli (2013) underlined that the violation of the classical cross-sectional independence assumption does not allow the assessment of valid statistical inference because the covariances of the SF errors can no longer be assumed to equal zero. Druska and Horrace (2004), Glass, Kenjegalieva, and Sickles (2013) and Han, Ryu, and Sickles (2016) are the first contributions taking spatial effects into consideration in SF models without making any distributional assumption on the inefficiency error term. Specifically, Druska and Horrace (2004) developed a spatial error model with time-invariant fixed effects, Glass, Kenjegalieva, and Sickles (2013) considered a fixed effect SAR model for panel data with time-variant technical efficiency and Han, Ryu, and Sickles (2016) proposed a SAR frontier model for panel data assuming a time-invariant inefficiency term.

Going beyond these first works taking into consideration spatial dependence in SF models without making any distributional assumption on the inefficiency error term  $u_i$ , the study by Adetutu et al. (2015) represents the first contribution introducing in SF models the spatial lag of the exogenous variables, making distributional assumptions on both the two error components. Glass, Kenjegalieva, and Sickles (2016) proposed a SAR stochastic frontier model and a spatial Durbin stochastic frontier model accounting for both global and local spatial dependence. Ramajo and Hewings (2018) further developed the spatial Durbin SF model by Glass, Kenjegalieva, and Sickles (2016) considering a time-varying decay specification for the inefficiency error term, as proposed by Battese and Coelli (1992) for non-spatial SF models. Gude, Alvarez, and Orea (2018) firstly introduced a generalized version of the spatial Durbin SF model incorporating time-varying exogenous influences on both the degree of global and local spatial spillovers for each time period. Finally, Tsukamoto (2019) extended the SAR stochastic frontier specification considering also the determinants of technical inefficiency, as proposed by Battese and Coelli (1995).

The second branch of literature on spatial SF models focused on introducing spatial dependence in the inefficiency term and/or in the random error term. Schmidt et al. (2009) is the first contribution in this field, making technical inefficiency

depend on a parameter that follows a prior distribution that captures unobserved spatial features. Differently from Schmidt et al. (2009), Areal, Balcombe, and Tiffin (2012) incorporated spatial dependence using an autoregressive specification of the inefficiency error term and estimated it using a Gibbs sampler and two Metropolis-Hastings steps. Similarly, Tsionas and Michaelides (2016) split the inefficiency error term into a spillover and an idiosyncratic component, developing a Bayesian estimator. Moreover, Fusco and Vidoli (2013) managed to solve the autoregressive error model of Areal, Balcombe, and Tiffin (2012) using a maximum likelihood estimation technique instead of a Bayesian approach. Moving from simple spatial models accounting for spatial dependence only in the inefficiency error term, the study by Herwartz and Strumann (2014) included both the SAR term and an autoregressive specification for the error term, obtaining a SARAR model. Orea and Alvarez (2019) introduced a SF model considering cross-sectional spatial effects both in the inefficiency and in the error term, capturing respectively behavioural and environmental spatial dependence. Using a two-step procedure, Skevas and Grashuis (2020) first employed a DEA approach to obtain the technical efficiency scores and then regressed the scores on firms' own characteristics and that of neighbours using a truncated regression model to assess the role of spatial spillovers on technical efficiency of US farmer cooperatives. Finally, Skevas and Skevas (2021) developed a generalized true random effects SF model with spatial terms in the inefficiency and the random error terms separating time-invariant from time-varying effects both in the noise and in the inefficiency component.

However, to our knowledge, none of the previous spatial stochastic frontier models allows evaluating the specific spillover effects resulting from each inefficiency determinant. Thus, our SDF-STE model firstly introduces spatial spillovers associated with the  $Z$  variables that locally influence all neighbouring firms.

### 3. The Model

#### 3.1. Model Specification

The specification of the SDF-STE model for panel data ( $N$  cross sectional units observed across  $T$  time periods) for  $i = 1, \dots, N$  and  $t = 1, \dots, T$  is defined as

$$Y_{it} = X_{it}\beta + \rho \sum_{j=1}^N w_{ij}Y_{jt} + \sum_{j=1}^N w_{ij}X_{jt}\theta + v_{it} - u_{it} \quad (1)$$

$$v_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v^2) \quad (2)$$

$$u_{it} \stackrel{iid}{\sim} \mathcal{N}^+(\mu_{it}, \sigma_u^2) \quad (3)$$

$$\mu_{it} = Z_{it}\phi + \sum_{j=1}^N w_{ij}Z_{jt}\delta, \quad (4)$$

where  $Y_{it}$  is the production output of firm  $i$  at time  $t$ ;  $X_{it}$  is a  $(1 \times k)$  vector containing the  $k$  production inputs used by firm  $i$  at time  $t$  with associated parameter vector  $\beta$  ( $k \times 1$ );  $\rho$  is the parameter that refers to the SAR term, capturing global spatial spillovers;  $w_{ij}$  is the generic element in the  $i$ -th row and  $j$ -th column of the block diagonal spatial weight matrix  $W$  ( $NT \times NT$ ) containing non-negative spatial weights to identify neighbours (indexed by  $j = 1, \dots, N$ ) and elements equal to zero on the main diagonal<sup>†</sup>;  $\theta$  is the parameter vector ( $k \times 1$ ) associated with the SLX term capturing exogenous local spatial spillovers. Finally, following the baseline specification for the error term as being made up of two components as proposed by Aigner, Lovell, and Schmidt (1977), the random error term  $v_{it}$  is assumed to be normally distributed with zero mean and variance  $\sigma_v^2$  and  $u_{it}$ , representing technical inefficiency, is distributed as a truncated normal random variable with known mean  $\mu_{it}$  and variance  $\sigma_u^2$ . Thus, in line with the half-normal specification for the inefficiency error term adopted, between the others, by Battese and Coelli (1995), Glass, Kenjegalieva, and Sickles (2016), and Tsukamoto (2019), we make technical inefficiency vary both across time and spatial units, accounting for time-varying technical inefficiency effects. Then, following the well-established specification for the inefficiency error term proposed by Battese and Coelli (1995) in their non-spatial SF production function for panel data, we model the mean  $\mu_{it}$  of the technical inefficiency term  $u_{it}$  as function of  $m$  exogenous determinants represented by the  $Z_{it}$  variables, with corresponding parameter vector  $\phi$  ( $m \times 1$ ). Moreover, we add in Eq.(4) the spatial lag of the determinants of technical inefficiency with associated parameter vector  $\delta$  ( $m \times 1$ ) capturing spatial dependence associated with the determinants of technical inefficiency of nearby firms. Through this new term, it is possible to capture all knowledge spillovers originating from nearby producers and affecting firms' efficiency levels.

Assumptions on Eq.(1)-(4) include (i)  $(I_{NT} - \rho W)$  non-singular, where  $I_{NT}$  is the  $(NT \times NT)$  identity matrix; (ii) row and columns sums of  $W$  and  $(I_{NT} - \rho W)^{-1}$ , before  $W$  is row-normalized, are uniformly bounded in absolute value as  $N$  goes to infinity (Kelejian and Prucha 1998, 1999). For a symmetric  $W$  the first assumption is always satisfied as long as the range of  $\rho$  is defined by  $\left(\frac{1}{\omega_{min}}, 1\right)$ , where  $\omega_{min}$  is the smallest real characteristic root of the spatial weight matrix  $W$  while the upper bound equals 1 for row-normalized  $W$ . Assumption (ii) limits the cross-sectional correlation, assuming that, when the distance separating two spatial units increases to infinity, it converges to zero. In particular, if  $W$  is a

<sup>†</sup>The  $W$  matrix is block diagonal because the locations of the  $N$  units are not assumed to change over time.



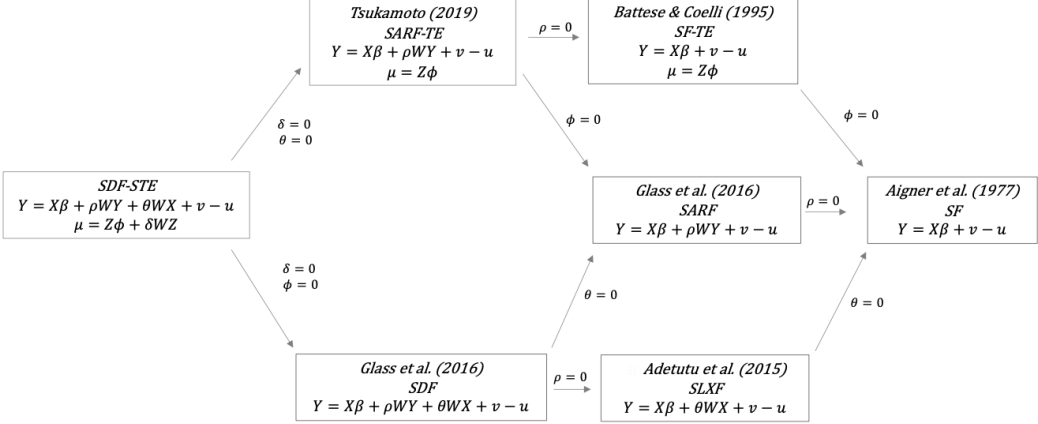


Figure 1: Nested Models

distance inverse spatial weight matrix, assumption (ii) can be guaranteed imposing a cut-off point  $d^*$  in  $W$  so that  $w_{ij} = 0$  if  $d_{ij} > d^*$ , while assumption (ii) is always satisfied if  $W$  is a binary contiguity matrix.

The novel SDF-STE model nests several existing spatial and non-spatial SF models, as represented in Figure 1. Imposing  $\delta = 0$  and  $\theta = 0$  our model reduces to the spatial autoregressive stochastic frontier model for panel data incorporating a model for technical inefficiency (SARF-TE) proposed by Tsukamoto (2019). If  $\delta = 0$  and  $\phi = 0$  our model becomes the spatial Durbin stochastic frontier model (SDF) estimated by Glass, Kenjegalieva, and Sickles (2016). Moreover, if  $\delta = 0$ ,  $\phi = 0$  and  $\theta = 0$  it coincides with the spatial autoregressive stochastic frontier model (SARF) (Glass, Kenjegalieva, and Sickles 2016). Imposing  $\delta = 0$ ,  $\rho = 0$  and  $\phi = 0$  our model is equivalent to the stochastic frontier introduced by Adetutu et al. (2015) that only includes the spatial lag of the exogenous variables (SLXF). Considering non-spatial SF models, if  $\delta = 0$ ,  $\theta = 0$  and  $\rho = 0$  our model reduces to the non-spatial stochastic frontier production function with a model for technical inefficiency (SF-TE) proposed by Battese and Coelli (1995). Finally, considering  $\delta = 0$ ,  $\theta = 0$ ,  $\rho = 0$  and  $\phi = 0$  our model becomes the classical non-spatial SF model by Aigner, Lovell, and Schmidt (1977). Therefore, following an approach similar to Manski (1993) in spatial econometric literature, our comprehensive specification allows for various parametric restrictions enabling a large set of modifications.

### 3.2. Estimation Procedure

Following Battese and Coelli (1995), the variance parameters are reparameterized as

$$\sigma^2 = \sigma_u^2 + \sigma_v^2, \quad \lambda = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}. \quad (5)$$

The loglikelihood function can be derived by substituting

$$\varepsilon_{it} = Y_{it} - X_{it}\beta - \rho \sum_{j=1}^N w_{ij} Y_{jt} - \sum_{j=1}^N w_{ij} X_{jt}\theta \quad (6)$$

$$\mu_{it} = Z_{it}\phi + \sum_{j=1}^N w_{ij} Z_{jt}\delta \quad (7)$$

in

$$\begin{aligned} \mathcal{L}(\Theta; y) = & \log |I_{NT} - \rho W| - \frac{NT}{2} (\log \sigma^2 + \log 2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T (\mu_{it} + \varepsilon_{it})^2 \\ & - \sum_{i=1}^N \sum_{t=1}^T \left[ \log \Phi \left( \frac{\mu_{it}}{\sigma \sqrt{\lambda}} \right) - \log \Phi \left( \frac{\mu_{it}(1-\lambda) - \varepsilon_{it}\lambda}{\sigma \sqrt{\lambda(1-\lambda)}} \right) \right], \end{aligned} \quad (8)$$

where,  $\Theta$  represents the vector of all parameters,  $\Phi$  is the cumulative distribution function of the standard normal random variable and  $\log |I_{NT} - \rho W|$  is the determinant of the Jacobian of the transformation from  $\varepsilon$  to  $Y_{it}$ . Indeed, since  $\frac{\partial \varepsilon}{\partial Y} = (I_{NT} - \rho W)$ , the endogeneity deriving from the inclusion of the spatial lag of the dependent variable has to be taken into account.

The unknown parameters can be simultaneously estimated numerically maximising the loglikelihood function in Eq.(8). In particular, we implemented a numerical constrained maximization algorithm in Matlab (codes are available under request). Since the parameter space for an autoregressive process is  $\left(\frac{1}{\omega_{min}}, 1\right)$ , where  $\omega_{min}$  is the smallest eigenvalue of  $W$ , the autoregressive parameter  $\rho$  should be bounded to the previous interval (LeSage and Pace 2009). Moreover,  $\sigma^2$  should be positive and  $0 \leq \lambda \leq 1$ . Specifically, if  $\lambda$  equals zero a simpler OLS should be preferred to the SF model because the variance of the inefficiency term is zero and therefore, the determinants of firms' efficiency can be included in the frontier function. Conversely,  $\lambda$  increases until 1 if the inefficiency effects are likely to be highly significant. Finally, to make the algorithm work better, the first derivatives of the loglikelihood function can be supplied to the program. The first derivatives of the likelihood function with respect to the unknown parameters are shown in Appendix A.

### 3.3. Marginal Effects

As evidenced by Elhorst (2014), in spatial models the  $\beta$  estimates cannot be interpreted as elasticities when the spatial lag of  $Y$  is included because changes in the generic regressor  $X_r$  of firm  $i$  influence the production output of firm  $j$ . Indeed, also for the SDF-STE model, the partial derivative of  $Y$  with respect to the generic regressor  $X_r$  ( $r = 1, \dots, k$ ) representing the marginal effect of  $X_r$  on  $Y$  does not clearly coincide with the estimated  $\beta_r$  coefficient, as shown by

$$\frac{\partial Y}{\partial X_r} = (I_{NT} - \rho W)^{-1}(I_{NT}\beta_r + W\theta_r). \quad (9)$$

LeSage and Pace (2009) proposed to compute the marginal effects of the independent variable  $X_r$  on  $Y$  starting from the matrix resulting from the right-hand side of Eq.(9) which depends on  $\rho$ ,  $\beta_r$  and  $\theta_r$ , representing respectively the spatial autoregressive parameter and the parameters associated with the generic regressor  $X_r$  and its spatial lag. Specifically, (i) the direct effect can be computed as the average of the diagonal elements of that matrix, (ii) the indirect effect is the average of the sum of the non-diagonal elements, (iii) the total effect equals the sum of the direct and the indirect effects.

Similarly to the  $\beta$  estimates, also the  $\phi$  estimates of the inefficiency model cannot be interpreted as elasticities due to the presence of the spatial lag of  $Y$ . In particular, the first derivatives of  $u$  with respect to a generic determinant  $Z_r$  with ( $r = 1, \dots, m$ ), is

$$\frac{\partial u}{\partial Z_r} = (I_{NT} - \rho W)^{-1}(I_{NT}\phi_r + W\delta_r). \quad (10)$$

Starting from the matrix obtained from the right-hand side of Eq.(10) which depends on  $\rho$  and on the parameters  $\phi_r$  and  $\delta_r$  related to the generic inefficiency determinant  $Z_r$  and its spatial lag, the marginal effects of  $Z_r$  on  $u$  can be computed straightforwardly, following a procedure analogous to the one described above. Thus, we adopt the method introduced by LeSage and Pace (2009) also to measure the marginal effects of the  $Z$  variables on the inefficiency level  $u$ . Accordingly, we measure (i) the direct effect of  $Z_r$  on  $u$  as the average of the diagonal elements of the matrix resulting from the product on the right-hand side of Eq.(10), (ii) the indirect effect as the average of the sum of the non-diagonal elements of the matrix resulting from Eq.(10), (iii) the total effect as the sum of the direct and the indirect effects.

Finally, to compute the standard errors or the t-values corresponding to the marginal effects of the  $X$  or the  $Z$  variables, two different methods can be used. First, as proposed by LeSage and Pace (2009), the distribution of the direct, indirect and total effects can be simulated starting from the variance-covariance matrix obtained from the maximum likelihood estimation procedure, and the standard errors can be approximated by the standard deviations associated with the different draws. Alternatively, following Glass, Kenjegalieva, and Sickles (2016), the delta method can be used.

### 3.4. Technical Efficiency Scores

Technical efficiency scores equal zero for fully inefficient firms and one for fully efficient firms and following Battese and Coelli (1988), they can be defined as  $TE_{it} = E(\exp(-u_{it})|\varepsilon_{it})$  for each firm  $i$  at each time period  $t$ . Concentrating on the SDF-STE model, TE scores taking productivity, input and determinants of firms' efficiency spillover effects into consideration can be calculated by substituting the parameter estimates in

$$TE_{it} = \exp \left[ -\mu_{it}(1-\lambda) + \varepsilon_{it}\lambda + \frac{1}{2} \left( \sigma \sqrt{(1-\lambda)\lambda} \right)^2 \right] \cdot \left\{ \frac{\Phi \left( \frac{\mu_{it}(1-\lambda) - \varepsilon_{it}\lambda}{\sigma \sqrt{(1-\lambda)\lambda}} - \sigma \sqrt{(1-\lambda)\lambda} \right)}{\Phi \left( \frac{\mu_{it}(1-\lambda) - \varepsilon_{it}\lambda}{\sigma \sqrt{(1-\lambda)\lambda}} \right)} \right\}. \quad (11)$$

## 4. Monte Carlo Simulations

To verify the finite sample properties of the estimated parameters, we run a Monte Carlo experiment simulating  $NT$  data. Moreover, for the first time in spatial SF literature, we also test the sensitivity of our spatial estimator to different choices of the spatial weight matrix.

To perform our simulation study we define the input variable  $X$  and the determinant of technical inefficiency  $Z$  as  $(NT \times 1)$  standard normal random vectors;  $W$  as a  $(NT \times NT)$  block diagonal and row standardized spatial weight matrix; the error term  $v$  as a  $(NT \times 1)$  normal random vector with zero mean and variance  $\sigma_v^2$  and the error term  $e$  as a  $(NT \times 1)$  truncated normal random vector with zero mean and variance  $\sigma_u^2$ , with point of truncation equal to  $-Z_{it}\phi - \sum_{j=1}^N w_{ij}Z_{jt}\delta$ , so that,  $e_{it} \geq -Z_{it}\phi - \sum_{j=1}^N w_{ij}Z_{jt}\delta$  for  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . The definition  $u = Z\phi + WZ\delta + e$  is consistent with specifying the inefficiency term  $u$  as a non-negative truncation of a normal distribution with mean  $\mu = Z\phi + WZ\delta$  and variance  $\sigma_u^2$  (Battese and Coelli 1995). Therefore, the dependent variable vector  $Y$  ( $NT \times 1$ ) is generated as

$$Y = (I_{NT} - \rho W)^{-1}(X\beta + WX\theta + v - Z\phi - WZ\delta - e). \quad (12)$$

To evaluate the impact of different choices on the estimation's bias, the standard deviation (SD) and the mean squared error (MSE), three blocks of simulations have been performed. In particular, in the first block of simulations, we evaluate how considering different values of  $N$  and  $T$  ( $N = 100, 200, 300$  and  $T = 5, 10, 15$ ), differently impact the estimated parameters. In this first block of simulations, the true values of the parameters are fixed at  $\{\beta = 0.50, \rho = 0.30, \theta = 0.30, \phi = 0.50, \delta = 0.50, \sigma_v^2 = 0.10, \sigma_u^2 = 0.10 (\sigma^2 = 0.20, \lambda = 0.50)\}$ . Conversely, in the second block of simulations, the true values of the parameters vary along

the different trials, while  $N$  and  $T$  are fixed at  $N = 100$  and  $T = 10$ . All the simulations are performed using 1000 repetitions. In these first two blocks of simulations, the spatial weight matrix  $W$  is defined as a binary contiguity matrix. In the third block of simulations, different kinds of spatial weight matrices have been considered (further details are described below). The results of the three blocks of simulations are shown in Table B1 of Appendix B for different values of  $N$  and  $T$ , in Table B2 of Appendix B for different true values of the parameters, and in Table 1 for different choices of the spatial weight matrix  $W$ .

The results of the first block of simulations show that the bias of all parameters is negligible, even considering small values for  $N$  and  $T$  ( $T = 5$  and  $N = 100$ ). This insight is in line with the simulation results of Carvalho (2018), finding that the performance of its spatial SF model including spatial dependence in the inefficiency component and accounting for unobserved heterogeneity is very good with small samples even for the spatial parameter. Nevertheless, we find that the bias reduces more and more as  $N$  and  $T$  increase, quickly approaching zero. Likewise, also the SD and the MSE tend to decrease increasing the numerosity of  $N$  and  $T$  and in particular, considering more time periods helps in obtaining a faster reduction of the bias and of the MSE. Moreover, as shown in Table B2, the estimates are robust to different changes in the true value of one parameter at a time, keeping all the others constant, even for a small sample size ( $N = 100$  and  $T = 5$ ). Indeed, the bias is always very near zero while the SD and the MSE remain fairly constant in all the simulations belonging to this second block.

Finally, Table 1 shows the results of the simulations considering different kinds of spatial weight matrices. If in the previous simulations a binary contiguity spatial weight matrix was taken into consideration, here  $W$  indicates an inverse distance spatial weight matrix,  $W50$  and  $W30$  indicate two inverse distance spatial weight matrices truncated at 50 and 30 km, respectively, and  $W250n$ ,  $W100n$  and  $W50n$  stand for three inverse distance spatial weight matrices considering only the 250, 100, and 50 nearest neighbours, respectively. All these spatial weight matrices are row standardized and they have been created starting from a random sub-sample of 300 observations belonging to the macro-area North-West of the AIDA sample used in the final application. Moreover, for all these simulations  $T$  is fixed at 5 and the number of replications is equal to 1000, as in the previous cases. The results of the simulation show that the parameters that do not depend on the spatial weight matrix  $W$  ( $\beta$ ,  $\phi$ , and the two variances  $\sigma^2$  and  $\lambda$ ) are not affected by changes in the spatial weight matrix. This result is in line with Areal, Balcombe, and Tiffin (2012) that, testing the sensitivity of the spatial parameter to different  $W$  matrices, showed that the model parameters that do not depend on  $W$  are not affected by different choices of the spatial weight matrix. In particular, we find that the bias, the SD and the MSE of  $\beta$  and  $\phi$  tend to be identical across the simulations using the same kind of  $W$ . On the contrary, in most cases, the bias, the SD and the MSE of  $\rho$ ,  $\theta$  and  $\delta$  (i.e. the parameters that depend on the spatial weight matrix) tend to decrease as the number of neighbours dimin-

Table 1: Sensitivity to the Choice of  $W$ 

	<b>W</b>			<b>W50</b>			<b>W30</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0007	0.0096	0.0001	0.0008	0.0098	0.0001	0.0008	0.0098	0.0001
$\rho$	-0.0091	0.0492	0.0026	-0.0056	0.0306	0.0010	-0.0037	0.0256	0.0007
$\theta$	0.0052	0.0534	0.0030	0.0043	0.0354	0.0013	0.0029	0.0299	0.0009
$\phi$	0.0002	0.0210	0.0004	0.0002	0.0204	0.0004	0.0002	0.0205	0.0004
$\delta$	0.0055	0.0878	0.0071	0.0045	0.0522	0.0027	0.0047	0.0439	0.0018
$\sigma^2$	0.0011	0.0138	0.0002	0.0002	0.0102	0.0001	-0.0003	0.0095	0.0001
$\lambda$	0.0069	0.0564	0.0034	0.0040	0.0429	0.0018	0.0017	0.0381	0.0015

	<b>W250n</b>			<b>W100n</b>			<b>W50n</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0002	0.0095	0.0001	0.0002	0.0096	0.0001	0.0002	0.0096	0.0001
$\rho$	-0.0031	0.0478	0.0026	-0.0020	0.0427	0.0020	-0.0016	0.0380	0.0015
$\theta$	-0.0005	0.0512	0.0027	-0.0010	0.0457	0.0021	-0.0010	0.0411	0.0017
$\phi$	-0.0012	0.0216	0.0004	-0.0012	0.0215	0.0004	-0.0012	0.0213	0.0004
$\delta$	-0.0032	0.0928	0.0083	-0.0028	0.0775	0.0059	-0.0021	0.0661	0.0044
$\sigma^2$	0.0004	0.0129	0.0002	0.0001	0.0120	0.0001	-0.0001	0.0114	0.0001
$\lambda$	0.0021	0.0516	0.0033	0.0013	0.0480	0.0027	0.0011	0.0448	0.0023

ishes, considering the same typology of  $W$ . Moreover, for  $N = 300$ , the binary contiguity spatial weight matrix used in the previous simulations results to be the one that minimizes the bias of  $\rho$ ,  $\theta$  and  $\delta$ , compared to the other specifications of  $W$ . However, the bias of the estimated parameters is still very near zero even considering different kinds of spatial weight matrices. Therefore, the choice of  $W$  has very little impact on the finite sample properties of our spatial estimator.

## 5. Application to the Italian Accommodation Sector

### 5.1. Data, Variables and Empirical Model

The SDF-STE model is estimated over a balanced sample of 4257 individual observations representing hotels located in Italy. Indeed, the Italian tourism and travel sector is one of the fastest growing and most profitable industrial sectors of Italy, contributing to 13% of the Italian GDP (including indirect effects) and giving employment to 14.7% of the Italian workforce in 2017 (OECD 2020). Moreover, the accommodation sector can be considered as highly characterized by spillover effects among nearby facilities. As highlighted by Novelli, Schmitz, and Spencer (2006), tourist destinations can be seen as forms of industrial clusters, made up of groups of small and medium enterprises (SMEs) that cooperate to build up a successful tourism product. SMEs clusters, through networking, alliances, active collaboration and innovation can succeed in successfully com-

peting in the global tourism market through local cooperation (Smeral 1998). Therefore, accommodation facilities located in tourism clusters can accumulate new knowledge and innovate more easily than isolated hotels (Marco-Lajara et al. 2016). Accordingly, estimating our model using a sample of Italian hotels should be an interesting case study providing noticeable insights into the role that spatial spillovers have in affecting the performance of Italian accommodation facilities. The data are collected from the AIDA Bureau Van Dijk database, which provides information on the consolidated accounts of Italian companies as well as on the geographical detail (i.e. longitudes and latitudes). In particular, we concentrated on the ATECO 55 sector over the period 2011–2019. Comparing the sample with the related population using census data from the Italian National Institute of Statistics (ISTAT) in 2018, our sample results to be fairly good in representing the ATECO 55 sector, covering the 7.85% of Italian hotels, the 26.88% of total employees and the 33.44% of the total value added generated in the sector. The empirical model defined based on a Cobb–Douglas production function for  $i, j = 1, \dots, N$  ( $i \neq j$ ) and  $t = 1, \dots, T$  is specified as

$$Y_{it} = \beta_0 + t\beta_t + t^2\beta_{2t} + L_{it}\beta_L + K_{it}\beta_K + \rho \sum_{j=1}^N w_{ij}Y_{jt} + \sum_{j=1}^N w_{ij}L_{jt}\theta_L \quad (13)$$

$$+ \sum_{j=1}^N w_{ij}K_{jt}\theta_K - u_{it} + v_{it},$$

$$\begin{aligned} \mu_{it} = & \phi_0 + \text{LP}_{it}\phi_{\text{LP}} + \text{OCF}_{it}\phi_{\text{OCF}} + \text{LEV}_{it}\phi_{\text{LEV}} + \text{NOD}_{it}\phi_{\text{NOD}} \\ & + \text{corp}_{it}\phi_{\text{corp}} + \text{ex}_{it}\phi_{\text{ex}} + \text{crisis}_{it}\phi_{\text{crisis}} + \sum_{j=1}^N w_{ij}\text{LP}_{jt}\delta_{\text{LP}} \\ & + \sum_{j=1}^N w_{ij}\text{OCF}_{jt}\delta_{\text{OCF}} + \sum_{j=1}^N w_{ij}\text{LEV}_{jt}\delta_{\text{LEV}} + \sum_{j=1}^N w_{ij}\text{NOD}_{jt}\delta_{\text{NOD}} \quad (14) \\ & + \sum_{j=1}^N w_{ij}\text{corp}_{jt}\delta_{\text{corp}} + \sum_{j=1}^N w_{ij}\text{ex}_{jt}\delta_{\text{ex}}. \end{aligned}$$

Specifically,  $Y_{it}$  represents the logarithm of the value added of hotel  $i$  at time  $t$  and  $L_{it}$  and  $K_{it}$  are the logarithms of the number of employees and of fixed assets respectively representing the two production inputs. Following Glass, Kenjegalieva, and Sickles (2016) we assume Hicks-neutral technical change and therefore the time trend variable  $t$  and its square  $t^2$  are added to the model specification. Moreover,  $\rho \sum_{j=1}^N w_{ij}Y_{jt}$  is the endogenous spatial lag capturing global productivity spillovers while  $\sum_{j=1}^N w_{ij}L_{jt}\theta_L$  and  $\sum_{j=1}^N w_{ij}K_{jt}\theta_K$  are the spatial lags related to the two input variables representing labour and capital local spillover effects.  $w_{ij}$  indicates the generic spatial weight belonging to the sparse row standardized

inverse distance spatial weight matrix  $W$  truncated at 50 km with all zeros on the main diagonal. Therefore, to capture spatial spillover effects, only those hotels located within a radius of 50 km are considered neighbours. Indeed, dense (i.e. not truncated) inverse distance spatial weight matrices take relations of firms with all territorial units into account, considering global spatial interactions without concentrating on local clusters (Kopczewska, Kudla, and Walczyk 2017). In the accommodation sector case, it is very unlikely that a hotel located in Sicily could affect a hotel located in Rome (unless they are in a corporate group). Therefore, it appears to be more reasonable to choose a cut-off point for  $W$  to evaluate the spatial effects only for hotels belonging to neighbouring destinations. Moreover, defining  $W$  using an inverse distance specification allows obtaining non-endogenous spatial weights. However, in subsection 5.4 we test the robustness of our results considering a dense  $W$  and a set of different truncation points for the distance function.

Table 2: Variables and Descriptive Statistics

Variable	Description	Measurement	Min	Mean	Max	SD
$Y$	Value Added	$\log(\text{value added})$	0.00	5.88	12.01	1.25
$L$	Labour	$\log(\text{number employees})$	0.00	2.18	7.47	1.10
$K$	Capital	$\log(\text{fixed capital})$	0.00	6.64	13.40	2.08
$t$	Time	1 for 2011, ..., 9 for 2019	1	5	9	2.58
LP	Labour Productivity	$\log(\text{revenues/staff costs})$	-9.27	1.36	7.27	0.62
LEV	Financial Leverage	$\log(\text{long debts/net assets})$	-9.54	0.01	7.80	1.63
OCF	Operational Cash Flow	$\log(\text{liquid assets/net assets})$	-12.81	-2.44	6.89	2.18
NOD	No Debts	1 if debts=0; 0 otherwise	0	0.22	1	0.42
ex	Experience	$\log(\text{years of activity})$	0.69	3.20	4.99	0.54
corp	N. Corporate Hotels	$\log(\text{number corporate})$	0.00	0.66	10.96	0.92
crisis	Crisis	1 for 2011, 2012, 2013	0	0.33	1	0.47

Considering the determinants of technical inefficiency, in the efficiency model a series of indicators measuring labour productivity (LP), operational cash flow (OCF) and financial leverage (LEV) are included, as well as information about the number of companies in corporate group (corp), the level of experience determined by the years of hotel's activity (ex) and two dummy variables identifying hotels with no long term debts (NOD) and the years of the crisis of the sovereign debt from 2011 to 2013 (crisis). Financial indicators such as operational cash flow and financial leverage are key elements to examine the variability in firms' performance because they are strongly associated with firms' failure and bankruptcy (Alberca and Parte 2020). In particular, higher values of OCF and LP are expected to positively influence firms' efficiency while LEV is expected to have an opposite effect. Accordingly, hotels that do not have long-term debts, hotels belonging to corporate groups and facilities with a major working experience in the sector are expected to reach an increased efficiency level. Further



details on the definition and measurement of all these variables and some brief descriptive statistics are given in Table 2. The efficiency model shown in Eq.(14) also includes the spatial lag of all the determinants of hotels' inefficiency with the exception of *crisis* to capture how hotels are influenced by the factors that determine the efficiency level of neighbouring accommodation facilities. Indeed, being surrounded by hotels belonging to corporate groups or by hotels with a high level of experience can positively influence neighbours according to knowledge spillover economic theory. Similarly, higher levels of LP and OCF, as well as lower levels of LEV, can be relevant indicators of more competitive destinations generating internal efficiency spillover effects. Specifically, combining  $\rho$  with the  $\phi$  and the  $\delta$  estimates, the indirect effects originating from the  $Z$  variables and affecting neighbouring firms' efficiency can be computed using the method proposed in Section 3.3.

Finally, as it can be noticed from Eq.(13)-(14), time-invariant heterogeneity is not captured by the model. Thus, the proposed model could be additionally extended in order to include also individual fixed or random effects. However, as underlined by Glass, Kenjegalieva, and Sickles (2016), the introduction of the spatial lag of the dependent variable measuring the degree of spatial dependence of  $Y$  in the cross-sections allows absorbing also spatial heterogeneity.

## 5.2. Estimation Results and Model Selection

Table 3 shows the estimation results of the SDF-STE model and of all the nested specifications. Specifically, SF denotes the basic non-spatial stochastic frontier model introduced by Aigner, Lovell, and Schmidt (1977), SF-TE stands for the non-spatial stochastic frontier production function with a model for technical inefficiency by Battese and Coelli (1995), SLXF (Adetutu et al. 2015), SARF and SDF (Glass, Kenjegalieva, and Sickles 2016) are the spatial stochastic frontier models introducing the spatial lag of the exogenous variables, of the dependent variable, and of both, respectively, while SARF-TE indicates the spatial stochastic frontier model by Tsukamoto (2019) including both the spatial lag of  $Y$  and a model for the determinants of firms' efficiency.

Comparing the estimation results of the nested models, it can be noticed that the  $\beta$  estimates and the  $\phi$  estimates are robust to the different specifications. Nevertheless, these coefficients cannot be interpreted in a meaningful way when the spatial lag of  $Y$  is included in the model because they no longer represent simple partial derivatives. Concentrating on the spatial autoregressive parameter  $\rho$ , it is evident that the estimates of  $\rho$  are all positive and significant at a 1% significance level indicating that positive spillover effects occur at the global level in the Italian accommodation sector. Moreover,  $\rho$  appears to be slightly overestimated if the determinants of firms' efficiency are not taken into account. Indeed,  $\rho$  equals 0.22 and 0.19 using respectively the SARF and the SARF-TE models while it equals 0.34 and 0.29 using the SDF and the SDF-STE specifications. Therefore, as already observed by Tsukamoto (2019) for the SARF and

the SARF-TE models, when the determinants of firms' efficiency are not considered in spatial stochastic frontier models,  $\rho$  absorbs some of the heterogeneity of technical inefficiency and tends to be overestimated.

Table 3: Nested Models: Estimation Results

	SF		SF-TE		SLXF		SARF		SDF		SARF-TE		SDF-STE	
	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
$\beta_0$	3.11***	53.79	6.27***	16.51	2.69***	44.52	1.88***	33.27	1.83***	36.53	4.47***	7.71	4.81***	32.27
$\beta_L$	0.75***	202.84	0.73***	201.47	0.74***	200.08	0.73***	192.87	0.74***	216.53	0.71***	203.17	0.72***	204.17
$\beta_K$	0.19***	91.86	0.23***	97.70	0.19***	86.86	0.19***	90.14	0.19***	90.00	0.22***	96.09	0.22***	96.17
$\beta_t$	0.09***	9.65	0.06***	6.85	0.10***	9.71	0.08***	10.89	0.07***	8.80	0.06***	6.82	0.05***	5.18
$\beta_{2t}$	-0.01***	-10.14	-0.01***	-8.00	-0.01***	-10.38	-0.01**	-10.29	-0.01***	-9.50	-0.01***	-8.43	-0.01***	-6.57
$\rho$	-	-	-	-	-	-	0.22***	21.17	0.34***	35.05	0.19***	18.82	0.29***	29.06
$\theta_L$	-	-	-	-	0.13***	9.39	-	-	-0.16***	-12.42	-	-	-0.16***	-12.59
$\theta_K$	-	-	-	-	0.02***	2.61	-	-	-0.05***	-7.34	-	-	-0.03***	-4.07
$\phi_0$	-	-	3.93***	10.33	-	-	-	-	-	-	3.16***	10.54	3.74***	12.07
$\phi_{LP}$	-	-	-0.08***	-14.98	-	-	-	-	-	-	-0.08***	-14.43	-0.08***	-12.07
$\phi_{OCF}$	-	-	-0.07***	-33.95	-	-	-	-	-	-	-0.06***	-34.44	-0.06***	-33.83
$\phi_{LEV}$	-	-	0.05***	22.90	-	-	-	-	-	-	0.05***	22.43	0.05***	22.10
$\phi_{NOD}$	-	-	-0.07***	-8.26	-	-	-	-	-	-	-0.07***	-7.88	-0.07***	-7.75
$\phi_{corp}$	-	-	-0.10***	-27.24	-	-	-	-	-	-	-0.09***	-25.53	-0.09***	-24.78
$\phi_{ex}$	-	-	-0.10***	-14.42	-	-	-	-	-	-	-0.08***	-12.74	-0.08***	-12.80
$\phi_{crisis}$	-	-	0.04**	2.33	-	-	-	-	-	-	0.03**	2.04	0.02	1.40
$\delta_{LP}$	-	-	-	-	-	-	-	-	-	-	-	-	-0.05***	-3.08
$\delta_{OCF}$	-	-	-	-	-	-	-	-	-	-	-	-	-0.02***	-3.50
$\delta_{LEV}$	-	-	-	-	-	-	-	-	-	-	-	-	0.02***	3.32
$\delta_{NOD}$	-	-	-	-	-	-	-	-	-	-	-	-	-0.01	-0.05
$\delta_{corp}$	-	-	-	-	-	-	-	-	-	-	-	-	-0.06***	-5.24
$\delta_{ex}$	-	-	-	-	-	-	-	-	-	-	-	-	-0.04**	2.25
$\sigma^2$	0.55	-	0.41	-	0.54	-	0.52	-	0.52	-	0.39	-	0.38	-
$\lambda$	0.35	-	0.40	-	0.34	-	0.34	-	0.34	-	0.24	-	0.49	-
$CI_L$	0.708	-	0.486	-	0.707	-	0.705	-	0.704	-	0.557	-	0.646	-
TE	0.798	-	0.607	-	0.797	-	0.796	-	0.795	-	0.668	-	0.748	-
$CI_U$	0.908	-	0.722	-	0.908	-	0.907	-	0.906	-	0.780	-	0.865	-

\*\*\* :  $pvalue \leq 0.01$ ; \*\* :  $pvalue \leq 0.05$ ; \* :  $pvalue \leq 0.10$

The estimated nested models are compared in Table 4. All the different indicators and tests considered highlight the strong preference for our spatial SF model with respect to previous specifications. First, the  $R^2$  indicates a better fit of our proposed spatial SF model to the data while considering the residual spatial correlation, our model is the one that minimizes the Moran's I statistic reaching a value of 0.025. However, even if the SDF-STE over-performs previous specifications in terms of residual spatial correlation, we cannot reject the null hypothesis of significant global spatial dependence in the residuals at a 10% significance level. This result suggests that another spatial structure related to the random error term may be required in order to consider unobserved but spatially correlated variables. Moreover, in Figure 2, we show the residuals' significance cluster map detecting the areas still characterized by significant local spatial dependence at 0.1%, 1%, and 5% significance levels. While highly significant spatial clusters are detected in the areas around Milan, Venice, and Rome, in Apulia, and along the northern Adriatic and Tyrrhenian coast using previous spatial and non-spatial SF models, our spatial specification is the one that most reduces the presence of local spatial dependence in the residuals. Indeed, only 39 observations out of 4257 still report significant spatial dependence at a 0.1% significance level with the SDF-STE model, while 408, 429, 506, 1012, 1162, and 1201 spatial units are significant with the SDF, SARF-TE, SARF, SLXF, SF-TE, and SF model, respectively. Other insights on the distribution of the residuals of all the nested models can be seen in Figure C1 of Appendix C. Finally, comparing the SDF-STE model with all the nested specifications using some likelihood ratio tests (LR), the null hypothesis assuming a better fit of the constrained model against the unconstrained one is always rejected in favour of our spatial estimator. Moreover, the SDF-STE is also the one that minimizes both the AIC and the BIC information criteria. In conclusion, in line with Areal and Pede (2021), which recommend using flexible spatial SF models which can take both frontier and efficiency spillovers into account, we find that our comprehensive model that considers three different sources of spatial dependence should be preferred to simpler specifications.

Table 4: Model Comparison

	$R^2$	Moran's I	LL	LR	p-value	AIC	BIC
SDF-STE	0.76	0.025*	-36137	-	-	72322	72384
SARF-TE	0.75	0.049***	-36168	63	0.00	72369	72410
SDF	0.67	0.027**	-37225	2175	0.00	74469	74495
SARF	0.66	0.056***	-37364	2453	0.00	74743	74764
SLXF	0.65	0.108***	-37835	3396	0.00	75688	75712
SF-TE	0.74	0.097***	-36691	1441	0.00	73413	73451
SF	0.64	0.112***	-37999	4055	0.00	76011	76029

\*\*\* :  $pvalue \leq 0.01$ ; \*\* :  $pvalue \leq 0.05$ ; \* :  $pvalue \leq 0.10$

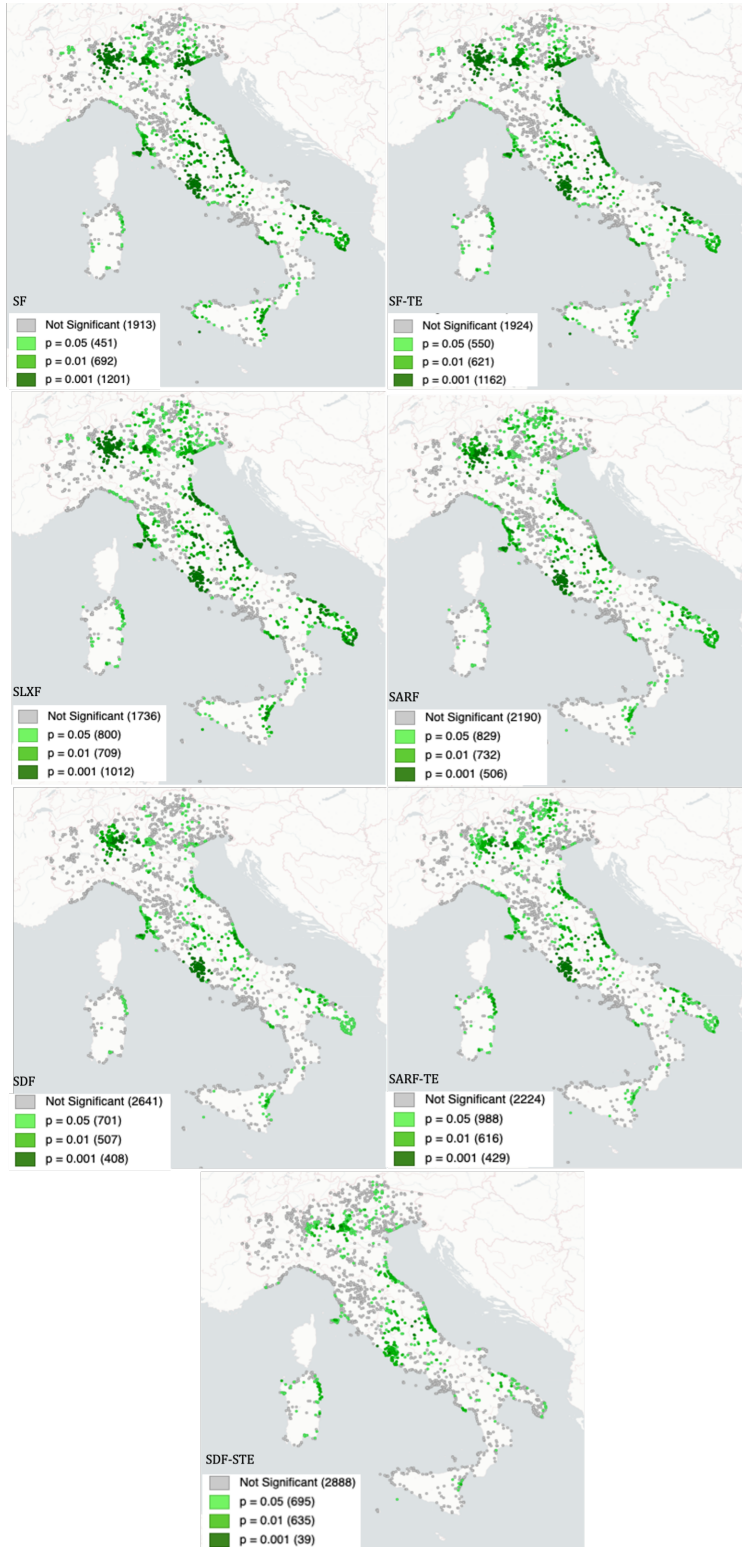


Figure 2: Significance Cluster Map: Residual Local Moran's I

### 5.3. *TE Scores and Marginal Effects*

The last three rows of Table 3 report the average technical efficiency scores with related 95% confidence intervals computed following Horrace and Schmidt (1996). Technical efficiency scores (TE) are largely overestimated using SF without a model for technical inefficiency. Indeed, the mean TE score is approximately equal to 0.80 for spatial and non-spatial SF without the inefficiency model while it equals 0.61 for the non-spatial SF-TE model and it reaches a value of 0.67 and 0.75 for the SARF-TE and for the SDF-STE models respectively. Therefore, both not including the Z variables and not considering spatial dependence in SF models lead to a severe bias in the estimation of TE scores.

Figure 3 shows the geographical distribution of the average TE scores per municipality from the SDF-STE model across the Italian territory for the year 2019. Considering the 7904 Italian municipalities from the 2018 classification, 2969 municipalities do not offer any accommodation facility (undefined category) according to the Industry and Services Census carried out by ISTAT in 2011, while 3718 of them are not covered in our sample. For the remaining 1217 municipalities, it can be noticed that hotels located in the Northern-East, in the Centre of Italy and in the Apulia region tend to be more efficient than others. In particular, hotels located both on the Tyrrhenian and on the Adriatic coast reach high efficiency scores, while the only internal areas achieving good efficiency levels are located in Trentino-Alto-Adige, Emilia-Romagna, Tuscany and Umbria.

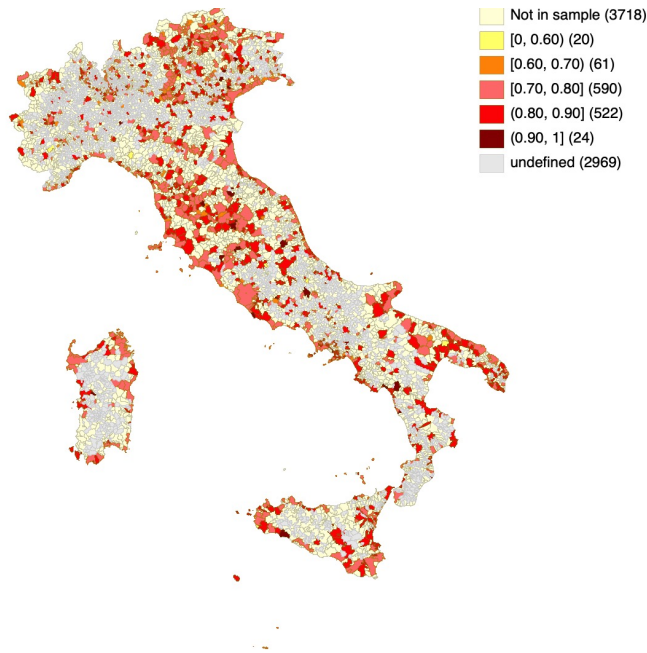


Figure 3: TE Scores across Italian Municipalities, year 2019

Table 5 shows the direct, indirect and total effects of the  $X$  variables on  $Y$  and of the  $Z$  variables on hotels' inefficiency level  $u$  with the related p-values and t-statistics computed using the delta method for the preferred SDF-STE model. Specifically, the SDF-STE model is the first one that allows capturing the specific indirect effects originating from each inefficiency determinant, providing noticeable insights to policymakers and destination managers on the specific spatial interactions occurring across neighbouring hotels.

Table 5: SDF-STE Model: Marginal Effects

	Direct		Indirect		Total	
	<i>Coeff.</i>	<i>t</i>	<i>Coeff.</i>	<i>t</i>	<i>Coeff.</i>	<i>t</i>
$L$	0.72***	204.57	0.07***	5.24	0.79***	54.78
$K$	0.22***	96.61	0.05***	4.85	0.27***	26.61
LP	-0.08***	-14.28	-0.10***	-4.39	-0.18***	-7.39
OCF	-0.06***	-34.39	-0.05***	-6.65	-0.11***	-13.78
LEV	0.05***	22.57	0.05***	5.43	0.10***	10.14
NOD	-0.07***	-7.72	-0.03	0.73	-0.09**	-2.30
corp	-0.09***	-25.47	-0.13***	-7.56	-0.22***	-12.65
ex	-0.09***	-12.89	-0.09***	-3.51	-0.18***	-6.32
crisis	0.02*	1.39	0.01*	1.39	0.03*	1.39

\*\*\* :  $pvalue \leq 0.01$ ; \*\* :  $pvalue \leq 0.05$ ; \* :  $pvalue \leq 0.10$

Focusing on the direct effects, it can be noticed that both labour and capital have a positive and significant direct effect on hotels' productivity. In particular, labour (0.72) has a larger direct effect compared to the one of capital (0.22), indicating that the Italian accommodation sector is a labour-intensive industry. Considering the determinants of hotels' inefficiency level, our results indicate that labour productivity (LP) has a negative direct effect on the inefficiency level of hotels as well as operational cash flow (OCF). On the contrary, making use of the financial leverage (LEV) has a positive direct effect on hotels' inefficiency, in line with Alberca and Parte (2020) for the Spanish accommodation industry. Likewise, hotels that do not have long-term debts are characterized by increased efficiency levels. As underlined by Orfila-Sintes, Crespi-Cladera, and Martinez-Ros (2005) and Arbelo, Arbelo-Pérez, and Pérez-Gómez (2018), hotels belonging to a chain tend to be more efficient than others being part of a favourable network that contributes to determining higher innovation rates. Moreover, also hotels' experience, proxied by the years of activity, positively contributes to increasing hotels' efficiency level. Finally, the crisis of the sovereign debt (captured through the variables *crisis*) contributed to increasing the inefficiency levels of Italian hotels in the period 2011-2013, as expected.

Considering the indirect effects, it can be noticed that both labour (0.07) and capital (0.05) have a strong and positive impact on neighbouring hotels. Thus, being surrounded by hotels that highly invest in labour force and capital endow-

ment contributes to strengthening the level of competitiveness of all neighbouring accommodation facilities. Similarly, positive spillover effects mainly arise from the determinants of firms' inefficiency levels. First, the indirect effects related to labour productivity (-0.10) and operational cash flow (-0.05) positively affect the efficiency level of neighbours. On the contrary, being close to hotels making wider use of the financial leverage generates negative spillover effects across nearby hotels (0.05). Moreover, locating close to hotels with high levels of experience (-0.09) or to corporate hotels (-0.13) has a negative influence on neighbouring hotels' inefficiency level, highlighting the existence of positive knowledge spillovers occurring across neighbouring Italian hotels.

In conclusion, this empirical application shows that while previous spatial SF models only capture spillover effects related to the frontier function, i.e. global productivity spillovers through  $\rho$  and spillovers related to the input variables through  $\theta_L$  and  $\theta_K$ , our novel spatial estimator allows considering local spillover effects associated with the inefficiency determinants through the  $\delta$  parameters. In particular, estimating the SDF-STE model over a sample of 4257 hotels located in Italy, it results that positive efficiency spillovers are related to labour productivity, operational cash flow, no debts, corporate groups, and hotels' experience while financial leverage generates negative feedbacks to neighbouring firms. All these insights would not have been available by estimating previous SF models. Between the spatial effects detected, we found that knowledge and labour spillovers are the most striking spatial effects occurring in the Italian accommodation sector. Indeed, locating near hotels with enhanced labour productivity generates positive spillover effects likely due to learning by observation, imitation, inter-firm exchanges and labour mobility (Hall and Williams 2008). This emulation tendency highlights, even more, the fundamental role that human resources have in the accommodation sector, as they are positively associated with customer satisfaction, service quality, competitive advantage and better organizational performance (Cho et al. 2006). Moreover, the diffusion of knowledge coming from more experienced hotels and from hotels in corporate groups benefiting from increased innovation rates positively affects neighbours. Thus, as detected by Hameed, Nisar, and Wu (2021), Stojčić, Vojvodić, and Butigan (2019), and Sundbo, Orfila-Sintes, and Sørensen (2007), the external environment in which hotels are embedded is a fundamental source of new knowledge and innovation. Our empirical findings suggest the importance for policymakers to exploit existing positive agglomeration economies in order to reinforce the cohesiveness and competitiveness of Italian accommodation facilities located in neighbouring destinations.

In future empirical applications of the proposed spatial SF model, Bayesian shrinkage and regularization can be considered to improve the statistical estimation of the SDF-STE model reaching an optimal trade-off between model complexity and out-of-sample performance (for a review of these methods see Polson and Sokolov (2019)). Moreover, in a further extension of this case study



it might be interesting to compare these results referring to the pre-pandemic period with the post-pandemic era using data from 2020, which are not yet available. Estimating the SDF-STE model over the two different sub-samples would allow analyzing the effect of the pandemic on spatial interactions occurring among neighbouring Italian accommodation facilities. In particular, on one hand, we expect reduced spatial dependence between neighbours due to social distancing and isolation, the reduction (or absence) of social events, decreased possibility to meet people due to mobility limitations, and the massive diffusion of digital communication technology for communications at the expense of in-person meetings. On the other hand, firms located in nearby areas may be more spatially dependent due to the localized dynamics of the pandemic and to specific policies and restrictions operating at the local level.

#### 5.4. Robustness Check

As usual in spatial models, the choice of the spatial scale considered to identify neighbours can considerably influence the modelled outcome because different spatial scales capture different spatial processes (Petrović, van Ham, and Manley 2021). Therefore, in this subsection, we provide further insights on how the estimated direct and indirect effects of the  $X$  and  $Z$  variables vary by choosing different truncation points for the inverse distance spatial weight matrix  $W$ . Specifically, to explore the effect of using different spatial scales, a dense inverse distance  $W$  (i.e. with no truncation point) is considered as well as matrices truncated at 200, 150, 100, 50, and 30km. As expected, the estimated direct effects represented in Figure C2 of Appendix C are not sensitive to the specification of  $W$  while the indirect effects in Figure 4 tend to decrease considering sparser spatial weight matrices. Indeed, while a dense  $W$  is useful to evaluate global spatial spillovers generated from everywhere on the map, considering a truncated  $W$  allows evaluating only local spillover effects. In this case, while the indirect effects related to NOD, OCF, L, K, and LEV appear to be fairly constant as the cut-off point of  $W$  changes, the magnitude of the spillover effects related to ex, corp, and LP clearly decreases as the radius choose to identify neighbours diminishes. Therefore, knowledge spillovers and labour productivity spillovers tend to cumulate and their intensity raises as the number of neighbours considered in the analysis increases confirming previous findings on the key role of these two kinds of spatial spillovers in the Italian accommodation sector. These insights are in line with the results of Cainelli and Ganau (2018), founding that the positive effects of agglomeration economies in the Italian manufacturing industry tend to raise with distance.

## 6. Concluding Remarks

Previous works on spatial SF models allowed evaluating the role of spatial proximity in affecting firms' productive outcomes by considering productivity and input

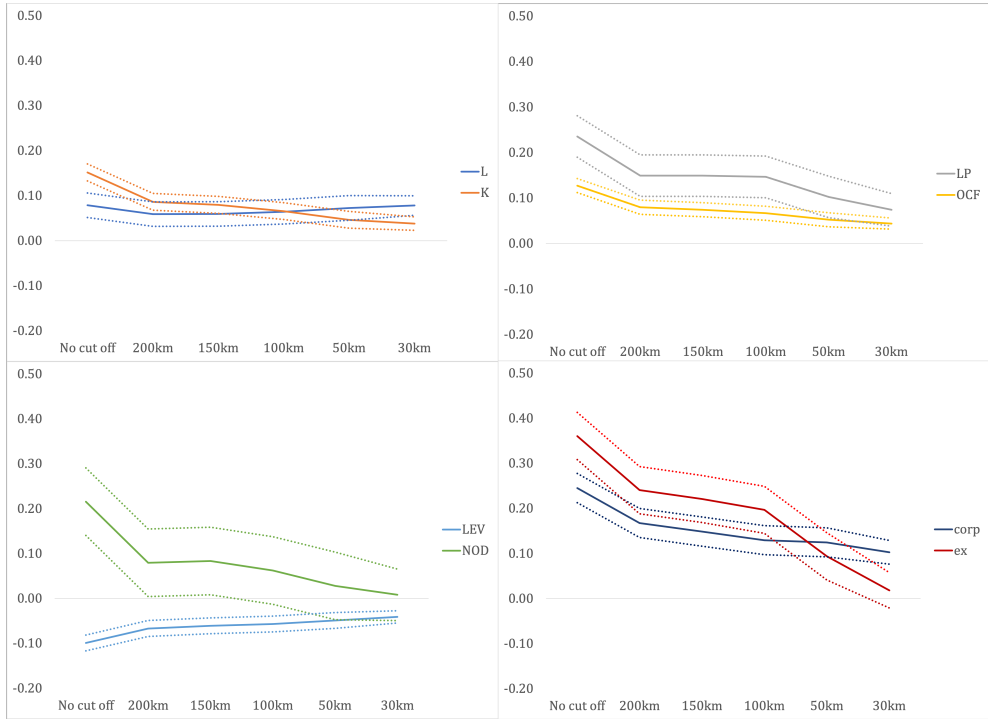


Figure 4: Indirect Effects for different truncation points of  $W$

Notes: For comparability, the marginal effects of the  $Z$  variables on the efficiency level  $-u$  are considered. Dashed lines: 95% confidence intervals.

spillovers or capturing the overall level of spatial dependence related to firms' inefficiency level. This paper extends the current literature on spatial stochastic frontier models introducing the possibility to consider specific spillover effects originating from the determinants of neighbouring firms' efficiency levels and affecting nearby producers. Indeed, as clearly assessed in economic geography literature, firms located in clusters tend to share abilities and new knowledge with neighbours allowing them to combine the production's factors more efficiently (Porter 1990).

Hence, the SDF-STE model developed in this paper takes three different kinds of spatial spillovers into consideration: (i) productivity, (ii) input and (iii) spillover effects in the determinants of firms' efficiency level identifying knowledge spillovers. Specifically, productivity spillovers globally affect all neighbouring units, while input and knowledge spillovers respectively influence the productive output and the efficiency level of nearby firms in a local way. The novel and most appealing feature of our spatial model concerns the possibility to evaluate the specific spatial effects arising from each inefficiency determinant allowing to precisely disentan-

gle different spillover effects originating from various inefficiency sources. The second interesting aspect of the SDF-STE model concerns its general and comprehensive specification that, other than detecting spillover effects related to the  $Z$  variables, it also captures global and local spatial spillovers associated with the frontier function. Finally, our new spatial estimator nests several existing spatial and non-spatial SF models allowing to select the model that best fits the data testing different restrictions through likelihood ratio tests.

The SDF-STE model can be estimated using maximum likelihood techniques. Performing different Monte Carlo simulations for alternative specifications of the spatial weight matrix, for the different true values of the parameters and for different sample sizes, it has been shown that the estimation results are unbiased even considering small samples. Moreover, besides the direct and indirect effects of the production inputs on  $Y$ , also the marginal effects related to the determinants of firms' efficiency on firms' inefficiency level  $u$  can be computed. Specifically, we follow the method proposed by LeSage and Pace (2009) to measure the direct and indirect effects of the input variables on  $Y$  and of the determinants of firms' efficiency on  $u$ . Finally, time-varying technical efficiency scores can be assessed for each unit following the approach proposed by Battese and Coelli (1988) and taking three different kinds of spatial spillover effects into consideration.

In the empirical application of this paper, a balanced panel of 4257 hotels located in Italy is considered to assess the role of spatial effects in influencing neighbouring hotels' productive outcomes. Results show that Italian hotels located in tourism clusters benefit from the existence of positive spatial spillovers resulting from spatial proximity while the only negative spillover effect is the one associated with the use of financial leverage, identifying less competitive and profitable destinations. In this case study, our spatial model better fits the data compared to the other spatial and non-spatial nested models resulting to be the preferred specification.

Our SDF-STE model could be further extended in order to include individual fixed or random effects. Moreover, besides evaluating spatial spillover effects associated with the mean of the inefficiency error term, also spatial dependence affecting the variance-covariance matrix of the inefficiency component can be considered. Finally, an additional spatial lag related to the random error term can be added to account for unobserved but spatially correlated variables.

## Figure legends

Figure 1: Nested Models

Figure 2: Significance Cluster Map: Residual Local Moran's  $I$

Figure 3: TE Scores across Italian Municipalities, year 2019

Figure 4: Indirect Effects for different truncation points of  $W$

## Funding Statement

The author received no financial support for the research, authorship, and publication of this article.

## Data Availability Statement

The data that support the findings of this study are available from <https://login.bvdinfo.com/R0/AidaNeo> but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. However, the codes used to perform the simulations will be shared by the author upon request.

## References

- Adams, J.D. and A.B. Jaffe (1996). "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data". In: *RAND Journal of Economics* 27.4, pp. 700–721.
- Adetutu, M., A.J. Glass, K. Kenjegalieva, and R.C. Sickles (2015). "The effects of efficiency and TFP growth on pollution in Europe: A multistage spatial analysis". In: *Journal of Productivity Analysis* 43, pp. 307–326.
- Aigner, D., C.K. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". In: *Journal of Econometrics* 6.1, pp. 21–37.
- Alberca, P. and L. Parte (2020). "Efficiency in the Holiday and Other Short-Stay Accommodation Industry". In: *Sustainability* 12, p. 9493.
- Arbelo, A., M. Arbelo-Pérez, and P. Pérez-Gómez (2018). "Estimation of Profit Efficiency in the Hotel Industry Using a Bayesian Stochastic Frontier Model". In: *Cornell Hospitality Quarterly* 59.4, pp. 364–375.
- Areal, F.J., K. Balcombe, and R. Tiffin (2012). "Integrating Spatial Dependence into Stochastic Frontier Analysis". In: *Agricultural and Resource Economics* 56.4, pp. 521–541.
- Areal, F.J. and V.O. Pede (2021). "Modelling Spatial Interaction in Stochastic Frontier Analysis". In: *Frontiers in Sustainable Food Systems* 5, p. 673039.
- Battese, G.E. and T.J. Coelli (1988). "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data". In: *Journal of Economics* 38.3, pp. 387–399.
- Battese, G.E. and T.J. Coelli (1992). "Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India". In: *Journal of Productivity Analysis* 3.1-2, pp. 153–169.
- Battese, G.E. and T.J. Coelli (1995). "A model for technical inefficiency effects in a stochastic frontier production function for panel data". In: *Empirical Economics* 20.2, pp. 325–332.

- Cainelli, G. and R. Ganau (2018). "Distance-based agglomeration externalities and neighbouring firms' characteristics". In: *Regional Studies* 52.7, pp. 922–933.
- Carvalho, A. (2018). "Efficiency spillovers in Bayesian stochastic frontier models: application to electricity distribution in New Zealand". In: *Spatial Economic Analysis* 13.2, pp. 171–190.
- Cho, S., R.H. Woods, S. Jang, and M. Erdem (2006). "Measuring the impact of human resource management practices on hospitality firms' performances". In: *Hospitality Management* 25, pp. 262–277.
- Coelli, T. (1995). "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis". In: *Journal of Productivity Analysis* 6, pp. 247–268.
- Druska, V. and W.C. Horrace (2004). "Generalized moments estimation for spatial panel data: Indonesian rice farming". In: *American Journal of Agricultural Economics* 86.1, 185–198.
- Elhorst, J.P. (2014). "MATLAB software for spatial panels". In: *International Regional Science Review* 37, 389–405.
- Fusco, E. and F. Vidoli (2013). "Spatial stochastic frontier models: controlling spatial global and local heterogeneity". In: *International Review of Applied Economics* 27.5, 679–694.
- Glass, A.J., K. Kenjegalieva, and R. Sickles (2013). "A Spatial Autoregressive Production Frontier Model for Panel Data: With an Application to European Countries". In: URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2227720](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2227720).
- Glass, A.J., K. Kenjegalieva, and R.C. Sickles (2016). "A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers". In: *Journal of Econometrics* 190.2, pp. 289–300.
- Griffith, R., R. Harrison, and J. van Reenen (2006). "How Special is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing". In: *American Economic Review* 96.5, pp. 1859–1875.
- Griliches, Z. (1992). "The Search for R&D Spill-Overs". In: *Scandinavian Journal of Economics* 94, pp. 29–47.
- Gude, A., I. Alvarez, and L. Orea (2018). "Heterogeneous spillovers among Spanish provinces: a generalized spatial stochastic frontier model". In: *Journal of Productivity Analysis* 50, pp. 155–173.
- Hall, C.M. and A.M. Williams (2008). *Tourism innovation*. London: Routledge.
- Hameed, W.U., Q.A. Nisar, and H. Wu (2021). "Relationships between external knowledge, internal innovation, firms' open innovation performance, service innovation and business performance in the Pakistani hotel industry". In: *International Journal of Hospitality Management* 92, p. 102745.
- Han, J., D. Ryu, and R.C. Sickles (2016). "How to measure spillover effects of public capital stock: a spatial autoregressive stochastic frontier model". In:

- Spatial Econometrics: Qualitative and Limited Dependent Variables*, Emerald Group Publishing Limited, pp. 259–294.
- Herwartz, H. and C. Strumann (2014). “Hospital Efficiency Under Prospective Reimbursement Schemes: an Empirical Assessment for the case of Germany”. In: *European Journal of Health Economics* 15, pp. 175–186.
- Hoover, E.M. (1948). *The Location of Economic Activity*. New York: McGraw-Hill.
- Horrace, W.C. and P. Schmidt (1996). “Confidence Statements for Efficiency Estimates from Stochastic Frontier Models”. In: *Journal of Productivity Analysis* 7, pp. 257–282.
- Kelejian, H.H. and I.R. Prucha (1998). “A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance”. In: *Journal of Real Estate Finance and Economics* 17, pp. 99–121.
- Kelejian, H.H. and I.R. Prucha (1999). “A generalized moments estimator for the autoregressive parameter in a spatial model”. In: *International Economic Review* 40.2, pp. 509–533.
- Kopczewska, K., J. Kudla, and K. Walczyk (2017). “Strategy of Spatial Panel Estimation: Spatial Spillovers Between Taxation and Economic Growth”. In: *Applied Spatial Analysis* 10, pp. 77–102.
- Lai, H. and C.J. Huang (2010). “Likelihood ratio tests for model selection of stochastic frontier models”. In: *Journal of Productivity Analysis* 34, pp. 3–13.
- LeSage, J.P. and R.K. Pace (2009). *Introduction to Spatial Econometrics*. Boca-Raton: Taylor & Francis.
- Levin, R.C. and P. Reiss (1988). “Cost-reducing and demand-creating R&D with spillovers”. In: *Rand Journal of Economics* 19, pp. 538–556.
- Manski, C.F. (1993). “Identification of endogenous social effects: the reflection problem”. In: *Review of Economic Studies* 60, pp. 531–542.
- Marco-Lajara, B., P.C. Zaragoza-Sáez, E. Claver-Cortés, and M. Úbeda García (2016). “Knowledge sources, agglomeration, and hotel performance”. In: *Journal of Business Research* 69, pp. 4856–4861.
- Meeusen, W. and J. van Den Broeck (1977). “Efficiency estimation from Cobb-Douglas production functions with composed error”. In: *International Economic Review* 18.2, pp. 435–444.
- Novelli, M., B. Schmitz, and T. Spencer (2006). “Networks, clusters and innovation in tourism: A UK experience”. In: *Tourism Management* 27, pp. 1141–1152.
- OECD (2020). *OECD Tourism Trends and Policies 2020*. Paris: OECD Publishing. DOI: <https://doi.org/10.1787/6b47b985-en..>
- Orea, L. and I.C. Alvarez (2019). “A new stochastic frontier model with cross-sectional effects in both noise and inefficiency terms”. In: *Journal of Econometrics* 213, pp. 556–577.

- Orfila-Sintes, F., R. Crespi-Cladera, and E. Martinez-Ros (2005). "Innovation activity in the hotel industry: Evidence from Balearic Islands". In: *Tourism Management* 26.6, pp. 851–865.
- Petrović, A., M. van Ham, and D. Manley (2021). "Where Do Neighborhood Effects End? Moving to Multiscale Spatial Contextual Effects". In: *Annals of the American Association of Geographers* 0, pp. 1–21.
- Polson, N.G. and V. Sokolov (2019). "Bayesian regularization: From Tikhonov to horseshoe". In: *WIREs Computational Statistics* 11.4, e1463.
- Porter, M. (1990). *The Competitive Advantage of Nations*. New York: Free Press.
- Porter, M. (1998). "Clusters and the New Economics of Competition". In: *Harvard Business Review*, pp. 77–90.
- Ramajo, J. and G.J. Hewings (2018). "Modelling regional productivity performance across Western Europe". In: *Regional Studies* 52.10, pp. 1372–1387.
- Schmidt, A.M., A.R.B. Moreira, S.M. Helfand, and T.C.O. Fonseca (2009). "Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency". In: *Journal of Productivity Analysis* 31, pp. 101–112.
- Skevas, I. and T. Skevas (2021). "A generalized true random-effects model with spatially autocorrelated persistent and transient inefficiency". In: *European Journal of Operational Research* 293, pp. 1131–1142.
- Skevas, T. and J. Grashuis (2020). "Technical efficiency and spatial spillovers: Evidence from grain marketing cooperatives in the US Midwest". In: *Agribusiness* 36, pp. 111–126.
- Smeral, E. (1998). "The Impact of globalization on small and medium enterprises: New challenges for tourism policies in European countries". In: *Tourism Management* 19.4, 371–380.
- Spence, M. (1984). "Cost Reduction, Competition, and Industry Performance". In: *Econometrica* 52, pp. 101–121.
- Stojčić, N., K. Vojvodić, and N. Butigan (2019). "Service Innovation and Firm Performance in the Hospitality Industry: Evidence from Tourism Driven Economy". In: *Tourism in Southern and Eastern Europe* 5, pp. 679–688.
- Sundbo, J., F. Orfila-Sintes, and F. Sørensen (2007). "The innovative behaviour of tourism firms—Comparative studies of Denmark and Spain". In: *Research Policy* 36, pp. 88–106.
- Syversen, C. (2011). "What determines productivity?" In: *Journal of Economic Literature* 49.2, pp. 326–365.
- Tsionas, E.G and P.G. Michaelides (2016). "A Spatial Stochastic Frontier Model with Spillovers: Evidence for Italian Regions". In: *Scottish Journal of Political Economy* 63.3, pp. 243–257.
- Tsukamoto, T. (2019). "A spatial autoregressive stochastic frontier model for panel data incorporating a model of technical inefficiency". In: *Japan and the World Economy* 50, pp. 66–77.

A Spatial Durbin Stochastic Frontier Model  
Introducing Spillover Effects in the Determinants of  
Firms' Efficiency

Supplementary Materials



## Appendix A - First Derivatives

Appendix A provides the partial derivatives of the loglikelihood function associated to the SDF-STE model with respect to the unknown parameters.

Defined  $m_{it}$  and  $s_{it}$  as in Eq.(A1),

$$m_{it} = \mu_{it}(1 - \lambda) - \varepsilon_{it}\lambda; \quad s_{it} = \sigma\sqrt{\lambda(1 - \lambda)} \quad (\text{A1})$$

the first derivatives of the loglikelihood function are shown in Eq.(A2)-Eq.(A8).

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{t=1}^T \sum_{i=1}^N \left( \frac{(\mu_{it} + \varepsilon_{it})}{\sigma^2} + \frac{\lambda \phi\left(\frac{m_{it}}{s_{it}}\right)}{s_{it} \Phi\left(\frac{m_{it}}{s_{it}}\right)} \right) X_{it} \quad (\text{A2})$$

$$\frac{\partial \mathcal{L}}{\partial \rho} = -T \text{tr} \left( (I_N - \rho W)^{-1} W \right) + \sum_{t=1}^T \sum_{i=1}^N \left( \frac{(\mu_{it} + \varepsilon_{it})}{\sigma^2} + \frac{\lambda \phi\left(\frac{m_{it}}{s_{it}}\right)}{s_{it} \Phi\left(\frac{m_{it}}{s_{it}}\right)} \right) \sum_{j=1}^N w_{ij} Y_{ij} \quad (\text{A3})$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \sum_{t=1}^T \sum_{i=1}^N \left( \frac{(\mu_{it} + \varepsilon_{it})}{\sigma^2} + \frac{\lambda \phi\left(\frac{m_{it}}{s_{it}}\right)}{s_{it} \Phi\left(\frac{m_{it}}{s_{it}}\right)} \right) \sum_{j=1}^N w_{ij} X_{ij} \quad (\text{A4})$$

$$\frac{\partial \mathcal{L}}{\partial \phi} = \sum_{t=1}^T \sum_{i=1}^N \left( -\frac{(\mu_{it} + \varepsilon_{it})}{\sigma^2} - \frac{\phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right) \sigma\sqrt{\lambda}} + \frac{\phi\left(\frac{m_{it}}{s_{it}}\right)(1 - \lambda)}{\Phi\left(\frac{m_{it}}{s_{it}}\right) s_{it}} \right) Z_{it} \quad (\text{A5})$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = \sum_{t=1}^T \sum_{i=1}^N \left( -\frac{(\mu_{it} + \varepsilon_{it})}{\sigma^2} - \frac{\phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right) \sigma\sqrt{\lambda}} + \frac{\phi\left(\frac{m_{it}}{s_{it}}\right)(1 - \lambda)}{\Phi\left(\frac{m_{it}}{s_{it}}\right) s_{it}} \right) \sum_{j=1}^N w_{ij} Z_{ij} \quad (\text{A6})$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{NT}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{t=1}^T \left( \frac{(\mu_{it} + \varepsilon_{it})^2}{\sigma^2} + \frac{\phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right) \sigma\sqrt{\lambda}} \frac{\mu_{it}}{\sigma\sqrt{\lambda}} - \frac{\phi\left(\frac{m_{it}}{s_{it}}\right)}{\Phi\left(\frac{m_{it}}{s_{it}}\right) s_{it}} \frac{m_{it}}{s_{it}} \right) \quad (\text{A7})$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{t=1}^T \sum_{i=1}^N \left( \frac{\phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma\sqrt{\lambda}}\right)} \frac{\mu_{it}}{2\sigma\lambda\sqrt{\lambda}} - \frac{\phi\left(\frac{m_{it}}{s_{it}}\right)}{\Phi\left(\frac{m_{it}}{s_{it}}\right)} \left( \frac{(\mu_{it} + \varepsilon_{it})}{s_{it}} + \frac{(1 - 2\lambda)\frac{m_{it}}{s_{it}}}{2(1 - \lambda)\lambda} \right) \right) \quad (\text{A8})$$

## Appendix B - Simulation Results

Table B1: Monte Carlo Simulation Results  
For different values of N and T

<b>T=5</b>	<b>N=100</b>			<b>N=200</b>			<b>N=300</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0011	0.0180	3.06e-04	0.0001	0.0123	1.55e-04	0.0004	0.0104	0.0001
$\rho$	-0.0013	0.0425	0.0017	-0.0002	0.0296	9.36e-04	-0.0010	0.0257	0.0007
$\theta$	0.0008	0.0371	0.0013	0.0006	0.0251	7.36e-04	0.0007	0.0231	0.0006
$\phi$	0.0004	0.0349	0.0012	-0.0002	0.0246	5.70e-04	-0.0005	0.0190	0.0004
$\delta$	0.0007	0.0576	0.0032	0.0020	0.0445	0.0021	0.0011	0.0373	0.0014
$\sigma^2$	-0.0023	0.0162	2.83e-04	-0.0001	0.0129	1.57e-04	-0.0002	0.0097	0.0001
$\lambda$	-0.0007	0.0656	0.0045	0.0022	0.0489	0.0025	-0.0001	0.0393	0.0017

<b>T=10</b>	<b>N=100</b>			<b>N=200</b>			<b>N=300</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0005	0.0126	0.0002	0.0010	0.0090	0.0001	-9.15e-05	0.0072	5.37e-05
$\rho$	-0.0008	0.0289	0.0009	-0.0015	0.0213	0.0005	-4.62e-04	0.0182	3.20e-04
$\theta$	-4.10e-05	0.0251	0.0006	0.0011	0.0184	0.0004	1.62e-04	0.0159	2.42e-04
$\phi$	0.0014	0.0246	0.0006	0.0010	0.0167	0.0003	4.71e-04	0.0134	1.79e-04
$\delta$	-0.0009	0.045	0.0022	0.0011	0.0332	0.0011	6.11e-04	0.0257	6.48e-04
$\sigma^2$	-0.0003	0.0120	0.0001	-0.0003	0.0090	0.0001	-1.52e-04	0.0068	5.06e-05
$\lambda$	0.0015	0.0443	0.0020	0.0009	0.0345	0.0012	3.06e-04	0.0276	7.92e-04

<b>T=15</b>	<b>N=100</b>			<b>N=200</b>			<b>N=300</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0004	0.0101	0.0001	-0.0010	0.0071	5.18e-05	2.30e-04	0.0070	3.39e-05
$\rho$	-0.0010	0.0233	0.0007	-0.0011	0.0183	3.28e-04	1.33e-04	0.0185	2.45e-04
$\theta$	0.0007	0.0204	0.0006	0.0012	0.0165	2.76e-04	-3.37e-04	0.0160	1.78e-04
$\phi$	-0.0005	0.0199	0.0004	0.0008	0.0134	1.88e-04	-2.42e-04	0.0137	1.28e-04
$\delta$	0.0011	0.0367	0.0014	0.0020	0.0263	7.13e-04	-9.41e-04	0.0281	5.19e-04
$\sigma^2$	-0.0002	0.0098	0.0001	4.81e-05	0.0072	5.30e-05	-5.67e-04	0.0067	3.29e-05
$\lambda$	-0.0001	0.0354	0.0017	0.0016	0.0309	6.25e-04	-7.17e-04	0.0309	5.13e-04

Table B2: Monte Carlo Simulation Results  
For different values of the parameters (N=100, T=10)

$\beta$	<b>0.15</b>			<b>0.65</b>			<b>0.90</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	-0.0001	0.0120	0.0002	0.0005	0.0127	0.0002	0.0004	0.0125	0.0001
$\rho$	0.0007	0.0302	0.0010	-0.0008	0.0286	0.0009	-0.0024	0.0314	0.0010
$\theta$	0.0005	0.0211	0.0005	0.0001	0.0282	0.0008	0.0016	0.0375	0.0014
$\phi$	0.0002	0.0253	0.0006	0.0014	0.0246	0.0006	0.0006	0.0265	0.0006
$\delta$	-0.0020	0.0451	0.0020	-0.0009	0.0449	0.0022	0.0033	0.0531	0.0026
$\sigma^2$	-0.0012	0.0120	0.0001	-0.0003	0.0119	0.0001	-1.83e-05	0.0132	0.0001
$\lambda$	-0.0025	0.0466	0.0024	0.0015	0.0440	0.0020	0.0018	0.0425	0.0025

$\rho$	<b>0.15</b>			<b>0.65</b>			<b>0.90</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	-0.0004	0.0119	0.0001	0.0001	0.0125	0.0001	-1.76e-04	0.0120	0.0001
$\rho$	-0.0014	0.0338	0.0011	-0.0011	0.0185	0.0004	-1.95e-04	0.0066	3.85e-05
$\theta$	0.0016	0.0273	0.0007	0.0009	0.0247	0.0006	2.34e-04	0.0240	0.0006
$\phi$	0.0009	0.0233	0.0005	-0.0009	0.0245	0.0006	1.93e-04	0.0246	0.0006
$\delta$	-0.0021	0.0453	0.0020	0.0033	0.0473	0.0020	2.59e-05	0.0453	0.0018
$\sigma^2$	-0.0009	0.0118	0.0001	-0.0001	0.0149	0.0002	-2.52e-04	0.0166	0.0002
$\lambda$	-0.0009	0.0461	0.0022	0.0022	0.0515	0.0029	-9.13e-04	0.0614	0.0039

$\theta$	<b>0.15</b>			<b>0.65</b>			<b>0.90</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	-0.0003	0.0123	0.0002	0.0001	0.0126	0.0001	0.0004	0.0135	0.0002
$\rho$	-0.0009	0.0338	0.0011	-3.09e-05	0.0297	0.0009	-0.0016	0.0259	0.0006
$\theta$	0.0001	0.0296	0.0009	-0.0001	0.0278	0.0009	0.0010	0.0275	0.0007
$\phi$	-0.0012	0.0232	0.0006	-0.0002	0.0233	0.0006	0.0005	0.0231	0.0006
$\delta$	0.0014	0.0507	0.0029	0.0011	0.0469	0.0025	0.0014	0.0433	0.0020
$\sigma^2$	-0.0002	0.0112	0.0001	-0.0012	0.0113	0.0001	-0.0008	0.0116	0.0001
$\lambda$	0.0007	0.0541	0.0026	-0.0009	0.0474	0.0022	0.0015	0.0485	0.0022

Table B2– continued from previous page

$\phi$	<b>0.15</b>			<b>0.65</b>			<b>0.90</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0006	0.0121	0.0001	-0.0003	0.0127	0.0002	0.0005	0.0136	0.0002
$\rho$	0.0008	0.0364	0.0011	-0.0001	0.0295	0.0009	-0.0001	0.0234	0.0006
$\theta$	0.0007	0.0299	0.0008	0.0002	0.0271	0.0007	-0.0004	0.0246	0.0006
$\phi$	0.0001	0.0273	0.0006	-0.0009	0.0243	0.0006	0.0016	0.0293	0.0009
$\delta$	0.0005	0.0439	0.002	0.0008	0.0502	0.0028	-0.0016	0.0526	0.0031
$\sigma^2$	-0.0015	0.0139	0.0001	-0.0011	0.0123	0.0002	-0.0006	0.0189	0.0004
$\lambda$	-0.0042	0.0509	0.0024	-0.0008	0.0503	0.0024	0.0004	0.0352	0.0014

$\delta$	<b>0.15</b>			<b>0.65</b>			<b>0.90</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	-2.93e-05	0.0124	0.0001	-3.21e-05	0.0138	0.0002	0.0004	0.0134	0.0002
$\rho$	9.77e-05	0.0311	0.0009	0.0009	0.0269	0.0007	-0.0001	0.0265	0.0008
$\theta$	5.69e-04	0.0277	0.0007	-0.0017	0.0293	0.0008	-0.0006	0.0251	0.0007
$\phi$	-5.83e-05	0.0245	0.0005	-0.0015	0.0292	0.0008	0.0015	0.0305	0.0010
$\delta$	4.98e-04	0.0457	0.0019	0.0004	0.0542	0.0027	-0.0012	0.0524	0.0028
$\sigma^2$	-7.56e-04	0.0136	0.0002	-0.0011	0.0191	0.0004	-0.0005	0.0181	0.0003
$\lambda$	9.08e-05	0.0457	0.0025	-0.0012	0.0378	0.0015	0.0005	0.0339	0.0013

$\sigma_v^2, \sigma_u^2$	<b>0.80, 0.10</b>			<b>0.10, 0.80</b>			<b>0.11, 0.09</b>		
	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>	<i>Bias</i>	<i>SD</i>	<i>MSE</i>
$\beta$	0.0006	0.0302	0.0009	-0.0001	0.0173	0.0003	-0.0003	0.0124	0.0001
$\rho$	-0.0016	0.0360	0.0014	0.0003	0.0233	0.0006	-0.0016	0.0317	0.0009
$\theta$	-0.0002	0.0511	0.0026	0.0015	0.0325	0.0010	0.0017	0.0279	0.0007
$\phi$	0.0014	0.0501	0.0027	0.0015	0.0506	0.0024	0.0010	0.0237	0.0005
$\delta$	-0.0020	0.0930	0.0095	-0.0061	0.0925	0.0080	-0.0018	0.0456	0.0020
$\sigma^2$	-0.0032	0.0428	0.0019	-0.0048	0.0552	0.0030	-0.0008	0.0122	0.0002
$\lambda$	0.0022	0.0343	0.0011	-0.0005	0.0161	0.0003	-0.0007	0.0495	0.0025

## Appendix C - Further Insights on the Empirical Application

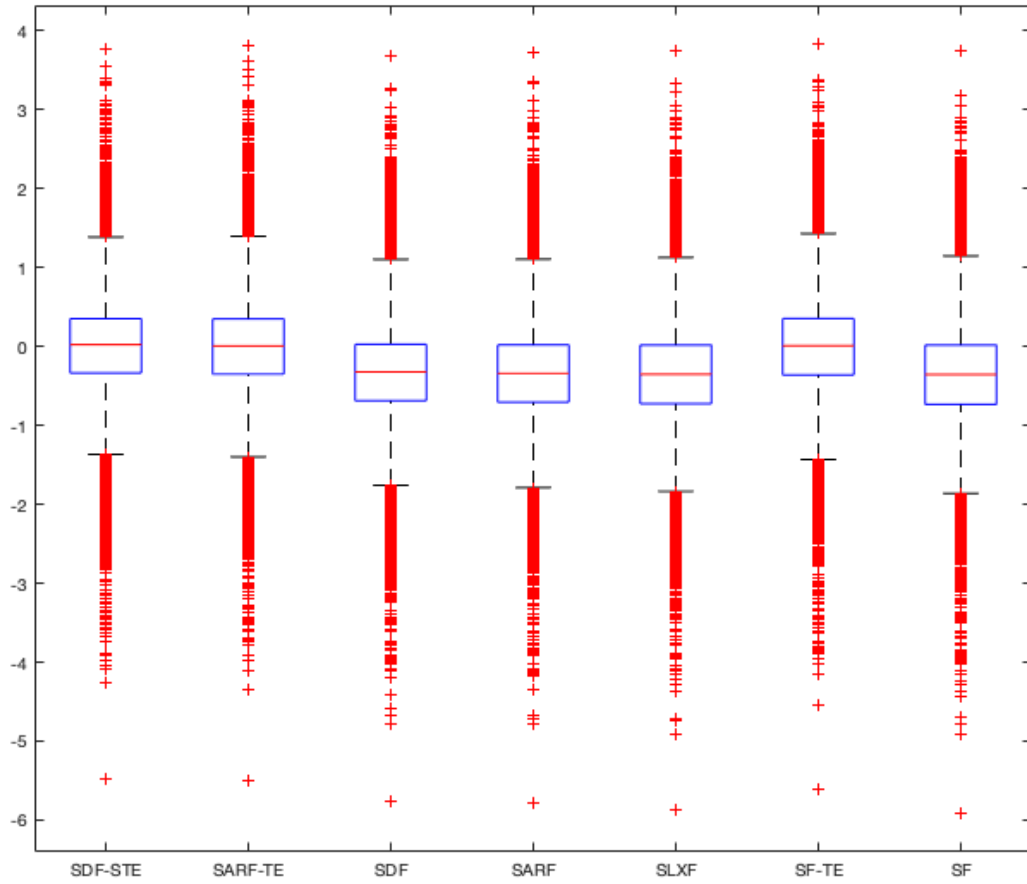


Figure C1: Nested Models: Residuals' Boxplots

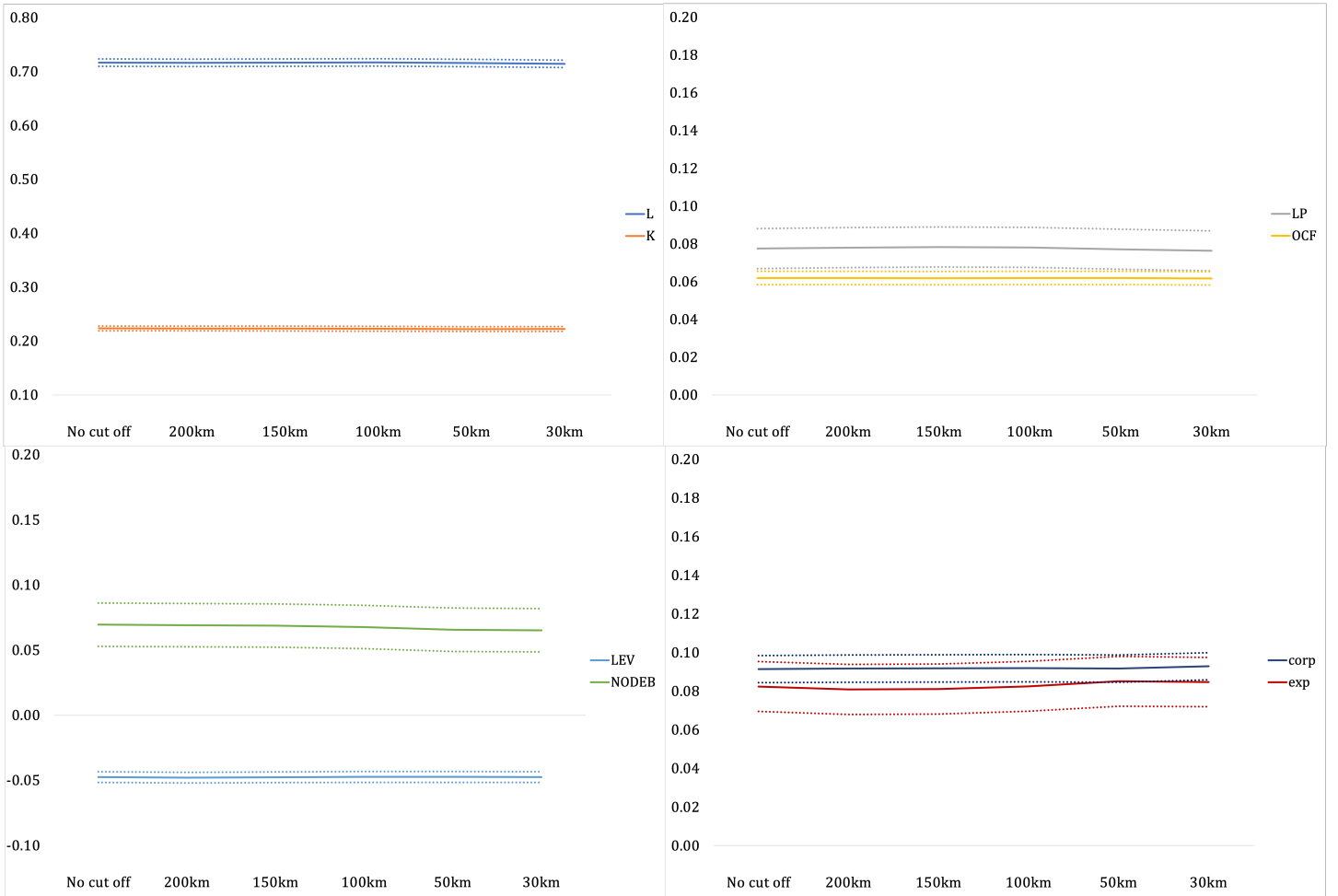


Figure C2: Direct effects for different truncation points of W

Dashed lines: 95% confidence intervals